Survey of data in sports

Maxime Cauchois

Stanford University

maxcauch@stanford.edu

April 3, 2019

Overview

1. Data is everywhere in sport...

2. How to make the most of it?

What use for data?

- Evaluate a player's performance: which ones should play?
- Predict a sport's outcome: what team is better than the other?
- Analyze the economy surrounding a team: what decisions should I make as a GM?
- Design an in-game strategy: who should shoot the ball, what play should I call, which point should I focus on?

Team data: a NBA case

			Net (Per 100 Poss)																				
Rk	Lineup	MP	FG	FGA	FG%	3P	3PA	3P%	eFG%	FT	FTA	FT%	PTS	ORB	ORB%	DRB	DRB%	TRB	TRB%	AST	STL	BLK	TOV	PF
1	T. Ariza | P. Beverley | J. Harden | D. Howard | D. Motiejunas	360:56	+1.8	-2.3	+.033	+5.3	+9.5	+.068	+.066	+1.8	+5.1	-.097	+10.7	+1.1	+3.7	+3.1	+3.7	+2.0	+4.7	+2.3	0.0	-2.4	-0.4	+0.8
2	T. Ariza | P. Beverley | J. Dorsey | J. Harden | D. Motiejunas	174:17	-0.5	+7.1	-.043	+3.6	+16.9	-.070	-.026	+3.0	+4.2	-.015	+5.6	+5.6	+9.4	-1.0	+9.4	+2.4	+5.6	+1.8	+1.2	-1.8	-1.4	+1.0
3	T. Ariza | P. Beverley | J. Harden | T. Jones | D. Motiejunas	156:50	-3.8	-7.2	-.004	+1.3	+7.2	-.048	+.007	+5.6	+6.9	-.004	-0.7	-2.7	-6.3	-3.2	-6.3	-3.0	-6.7	-1.1	-0.3	-1.3	+0.9	-1.8
4	T. Ariza | P. Beverley | J. Harden | D. Motiejunas | J. Smith	99:57	-5.6	-0.5	-.060	+2.6	+8.7	-.018	-.045	+4.6	+5.6	+.007	-4.1	0.0	-5.9	-9.7	-5.9	-4.8	-11.8	-3.6	+1.5	+3.6	-3.6	-0.3
5	T. Ariza | P. Beverley | J. Harden | D. Howard | J. Smith	92:17	-8.5	-2.4	-.085	+1.9	+14.0	-.122	-.073	-2.3	+0.9	-.126	-17.4	-0.4	-5.0	-6.2	-5.0	-3.6	-8.2	-8.0	+0.9	-3.7	+2.7	+1.1
6	T. Ariza | J. Harden | D. Howard | T. Jones | J. Terry	84:33	+0.5	-10.7	+.066	+7.6	+10.8	+.173	+.115	+8.3	+7.8	+.083	+16.9	-4.2	-4.4	+5.2	-4.4	+0.6	+1.4	+3.4	+1.7	+1.2	+3.9	0.0
7	C. Brewer | J. Harden | T. Jones | J. Smith | J. Terry	76:45	+4.6	-12.1	+.113	+4.6	-1.1	+.170	+.145	+7.9	+7.8	+.082	+21.7	-7.8	-9.9	+6.5	-9.9	-0.9	-1.9	+10.2	+2.5	+4.8	+0.9	-3.5
8	T. Ariza | T. Black | J. Harden | D. Motiejunas | J. Terry	71:20	+6.9	-1.6	+.086	+7.9	+8.8	+.151	+.131	+2.3	+1.0	+.083	+23.9	+3.7	+12.8	+11.4	+12.8	+7.7	+16.7	+2.4	-5.0	-4.3	+6.5	+1.7
9	T. Ariza | C. Brewer | D. Motiejunas | J. Smith | J. Terry	57:25	+2.0	-2.0	+.031	+3.9	+7.1	+.043	+.054	+2.1	+4.0	-.045	+9.9	-4.2	-7.8	-2.8	-7.8	-3.2	-6.4	+11.5	+1.6	+2.4	-6.7	+3.2
10	T. Ariza | I. Canaan | J. Harden | D. Howard | D. Motiejunas	55:48	+3.3	-14.7	+.108	+4.5	+2.5	+.141	+.142	+9.1	+13.6	+.037	+20.3	-3.8	-4.4	+2.2	-4.4	-0.9	-1.9	-4.8	-0.8	+7.5	+2.6	-2.3
11	T. Ariza | C. Brewer | T. Jones | J. Smith | J. Terry	53:29	+3.7	-5.5	+.061	+8.3	+17.4	+.089	+.111	-0.9	+5.5	-.250	+14.7	-6.4	-10.2	-2.8	-10.2	-4.6	-8.5	+4.6	-0.9	+4.6	-0.9	-2.3
12	T. Ariza | P. Beverley | T. Black | J. Harden | D. Motiejunas	51:09	-0.8	-5.9	+.020	+8.0	+8.5	+.167	+.069	-5.5	-8.6	+.082	+0.9	-3.3	-4.3	+1.7	-4.3	-1.0	-2.3	-2.5	-1.8	-1.0	+1.7	+5.6
13	C. Brewer | J. Harden | D. Motiejunas | J. Smith | J. Terry	50:40	+9.4	+2.1	+.103	+5.8	+9.5	+.091	+.137	+3.7	+7.5	-.071	+28.2	-3.0	-6.6	-1.7	-6.6	-2.4	-6.0	+7.6	+6.9	-3.9	-10.9	-0.5
14	C. Brewer | J. Dorsey | J. Harden | P. Prigioni | J. Smith	45:29	+0.4	+8.0	-.035	-0.9	+7.4	-.135	-.045	+13.4	+26.7	-.145	+13.3	+14.4	+19.9	-0.3	+19.9	+7.1	+14.6	-0.8	+5.0	-4.1	-3.9	-7.6
15	T. Ariza | J. Harden | D. Motiejunas | K. Papanikolaou | J. Terry	42:49	+1.1	-1.9	+.021	+6.7	+21.8	+.099	+.060	-0.7	-8.9	+.246	+8.2	-0.4	+0.5	+2.5	+0.5	+0.6	+1.1	+6.2	-4.3	-2.1	+0.7	-1.7
16	T. Ariza | C. Brewer | D. Howard | P. Prigioni | J. Smith	41:30	+20.0	+15.0	+.146	+1.3	+6.3	-.027	+.147	-1.3	+3.8	-.215	+40.0	+7.5	+15.8	+8.8	+15.8	+8.1	+16.5	+11.3	+3.8	+1.3	-8.8	-1.9
17	T. Ariza | J. Harden | D. Howard | D. Motiejunas | J. Terry	41:07	+8.6	+6.1	+.076	+2.6	+4.6	+.065	+.090	-11.9	-3.3	-.324	+7.9	+0.3	+2.6	+5.3	+2.6	+3.5	+8.1	+10.4	+1.9	-0.4	+2.0	-0.6
18	T. Ariza | P. Beverley | J. Harden | T. Jones | J. Smith	40:45	+1.9	-4.1	+.040	+8.2	+13.7	+.121	+.087	-3.8	-5.0	-.013	+8.2	-14.5	-22.4	-5.9	-22.4	-10.1	-19.0	+0.3	+2.6	+8.0	-2.5	+4.0
19	P. Beverley | C. Brewer | J. Harden | D. Motiejunas | J. Smith	38:38	+2.8	-12.8	+.099	-3.8	-11.5	-.006	+.084	-1.7	-4.2	+.054	+0.1	-5.0	-6.5	+8.2	-6.5	+1.2	+2.6	-1.7	-7.0	+3.7	+9.2	+1.2
20	T. Ariza | J. Harden | D. Motiejunas | J. Smith | J. Terry	37:46	+4.8	+18.1	-.036	+7.2	+18.7	+.014	-.012	-8.2	-9.5	-.035	+8.6	+5.8	+1.7	-8.8	+1.7	-1.4	-2.9	-1.0	+11.1	0.0	-6.8	-0.7
	Team Average	**3961:00**	**-0.7**	**-1.9**	**+.001**	**+4.1**	**+10.1**	**+.026**	**+.026**	**+0.9**	**+2.5**	**-.036**	**+3.6**	**-0.2**	**-0.3**	**0.0**	**-0.3**	**-0.1**	**-0.2**	**+0.9**	**+0.6**	**-0.4**	**+0.3**	**+0.5**

Figure: NBA performance in function of the squad on the field

A look at individual data

#	PLAYER	GP	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	STL	BLK	TOV	EFF
1	James Harden	38	37.1	34.1	9.9	22.6	43.8	4.9	12.6	39.0	9.4	11.0	85.6	0.7	5.4	6.1	8.7	2.0	0.6	5.7	31.6
2	Anthony Davis	37	37.0	28.9	10.4	20.2	51.4	0.9	2.9	32.1	7.2	9.0	80.4	3.4	9.9	13.2	4.4	1.8	2.7	2.0	37.5
3	Stephen Curry	31	34.5	28.8	9.6	19.8	48.7	5.0	11.4	44.3	4.5	4.9	91.5	0.7	4.5	5.2	5.6	1.3	0.3	3.1	27.5
4	Kevin Durant	42	35.6	28.2	9.7	19.2	50.7	1.8	5.0	36.8	7.0	7.6	91.3	0.5	6.9	7.3	6.1	0.8	1.0	3.4	30.0
5	LeBron James	34	34.6	27.3	10.0	19.3	51.8	2.0	5.6	35.6	5.3	7.8	68.2	0.9	7.4	8.3	7.1	1.3	0.7	3.4	29.6
6	Kawhi Leonard	34	34.6	27.1	9.5	18.9	50.2	1.7	4.7	36.9	6.4	7.6	84.5	1.3	6.6	7.9	3.1	1.9	0.5	2.0	27.9
7	Joel Embiid	40	33.7	26.9	8.9	18.2	49.0	1.2	3.8	30.1	8.0	9.9	80.6	2.3	11.1	13.5	3.5	0.6	1.9	3.5	31.6
8	Paul George	40	35.7	26.9	9.2	20.6	44.8	3.4	8.8	38.5	5.1	6.2	82.2	1.5	6.5	8.0	3.9	2.2	0.6	2.7	26.4
9	Giannis Antetokounmpo	38	33.6	26.6	9.9	17.1	58.1	0.4	2.4	16.7	6.3	9.0	69.9	2.4	10.4	12.8	6.0	1.3	1.5	4.1	34.2
10	Damian Lillard	43	35.2	25.8	8.5	19.0	44.6	2.9	7.6	38.4	5.9	6.6	89.8	0.8	3.7	4.5	6.1	1.0	0.5	2.8	24.0
11	Blake Griffin	39	36.1	25.6	8.7	18.1	47.8	2.4	6.6	36.3	5.8	7.7	75.7	1.3	7.0	8.3	5.3	0.8	0.5	3.8	25.3
12	Kemba Walker	41	34.3	24.9	8.5	19.7	43.3	3.2	8.9	35.8	4.7	5.7	81.6	0.5	3.8	4.3	5.7	1.2	0.5	2.5	21.9
13	Devin Booker	31	35.2	24.8	8.8	19.4	45.2	2.3	6.9	32.6	5.0	6.0	84.3	0.6	3.1	3.7	6.9	1.1	0.2	3.8	21.3
14	Bradley Beal	43	36.7	24.4	9.1	19.2	47.3	2.5	7.1	35.2	3.7	4.7	80.0	1.0	3.8	4.8	4.8	1.3	0.8	2.7	22.5
15	Zach LaVine	36	34.3	23.6	8.4	18.1	46.1	1.8	5.0	36.1	5.1	5.9	86.8	0.4	3.9	4.4	4.2	0.9	0.4	3.6	19.4
16	Kyrie Irving	37	32.1	22.6	8.7	17.7	49.0	2.6	6.3	40.6	2.7	3.2	85.5	1.2	3.6	4.8	6.4	1.6	0.4	2.5	23.9
17	Karl-Anthony Towns	42	33.5	22.3	8.1	16.3	49.9	1.7	4.5	37.9	4.3	5.2	83.5	3.3	8.9	12.2	3.0	0.9	1.9	2.9	28.4
18	DeMar DeRozan	43	35.3	22.2	8.5	18.1	46.8	0.2	0.9	17.5	5.1	6.1	82.9	0.7	5.4	6.2	6.5	1.0	0.5	2.5	23.2
19	Klay Thompson	42	34.5	21.8	8.5	18.8	45.3	2.8	7.5	36.6	2.0	2.5	80.2	0.5	3.7	4.1	2.0	1.2	0.7	1.6	17.5
20	Russell Westbrook	33	35.3	21.3	8.4	19.8	42.3	1.1	4.8	22.9	3.4	5.5	62.1	1.3	9.5	10.8	10.7	2.5	0.3	4.7	27.5

Figure: NBA 2018-2019 Individual Performance

Major Question

Do they tell you everything of a player's actual performance?

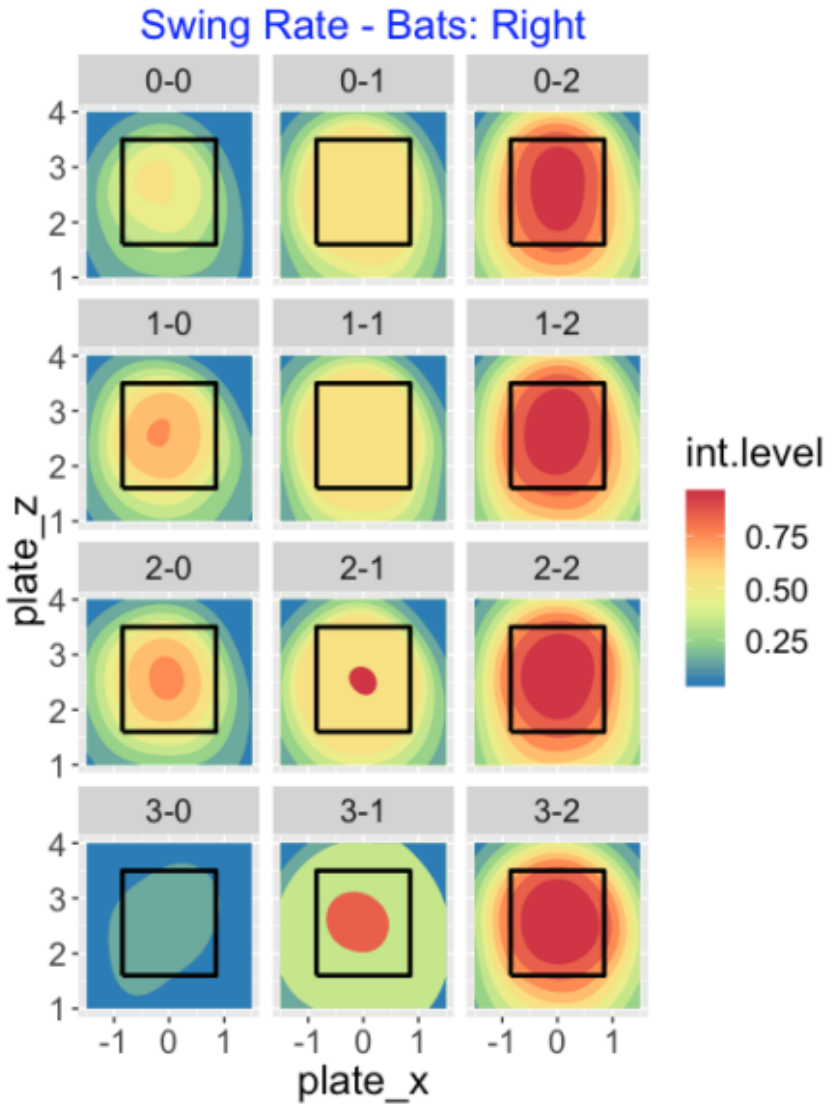
In-game strategy with data

A1 | streak

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	team_sho	opp_shot	team_q_s	opp_q_sc	diff_shot	diff_q_sc	diff_gain	clock_raw	clock	shot_clocl	shot_clocl	clock_star	type	distance	made	wins	away	losses	team	quarter	gametime	link	player	season
2	47	52	49	54	-5	-5	0	29	29	5	5	34	3-pt	24	FALSE	1	0	0	POR	End of 2nc	10:30p	http://ww	S. Blake	2015
3	75	77	75	77	-2	-2	0	26	26	11	11	37	2-pt	3	FALSE	1	0	0	POR	End of 3rd	10:30p	http://ww	C. Kaman	2015
4	103	89	106	89	14	17	3	33	33	9	9	42	2-pt	7	FALSE	1	0	0	POR	End of 4th	10:30p	http://ww	C. McColl	2015
5	46	48	48	48	-2	0	2	29	29	6	6	35	2-pt	1	TRUE	1	1	1	POR	End of 2nc	10:00p	http://ww	C. Kaman	2015
6	69	71	71	71	-2	0	2	30	30	13	13	43	2-pt	2	TRUE	1	1	1	POR	End of 3rd	10:00p	http://ww	C. Kaman	2015
7	91	99	94	103	-8	-9	-1	33	33	18	18	51	3-pt	23	FALSE	1	1	1	POR	End of 4th	10:00p	http://ww	L. Aldridge	2015
8	68	68	68	68	0	0	0	33	33	0	0	33	3-pt	23	FALSE	1	0	2	POR	End of 3rd	9:00p	http://ww	S. Blake	2015
9	90	88	90	95	2	-5	-7	33	33	11	11	44	3-pt	24	FALSE	1	0	2	POR	End of 4th	9:00p	http://ww	L. Aldridge	2015
10	99	80	101	82	19	19	0	30	30	21	21	51	2-pt	0	TRUE	2	0	2	POR	End of 4th	10:00p	http://ww	T. Robinsc	2015
11	42	48	46	50	-6	-4	2	34	34	4	4	38	2-pt	1	TRUE	3	0	2	POR	End of 2nc	10:30p	http://ww	D. Lillard	2015
12	38	31	38	33	7	5	-2	31	31	7	7	38	2-pt	6	FALSE	3	1	3	POR	End of 1st	3:30p	http://ww	C. McColl	2015
13	100	104	102	106	-4	-4	0	35	35	20	20	55	2-pt	2	TRUE	3	1	3	POR	End of 4th	3:30p	http://ww	D. Lillard	2015
14	59	54	64	56	5	8	3	34	34	15	15	49	2-pt	21	TRUE	4	0	3	POR	End of 2nc	9:00p	http://ww	L. Aldridge	2015
15	19	35	21	35	-16	-14	2	34	34	15	15	49	3-pt	24	FALSE	5	0	3	POR	End of 1st	10:00p	http://ww	S. Blake	2015
16	46	62	51	64	-16	-13	3	36	36	9	9	45	3-pt	25	TRUE	5	0	3	POR	End of 2nc	10:00p	http://ww	D. Lillard	2015
17	75	85	75	85	-10	-10	0	28	28	9	9	37	2-pt	20	FALSE	5	0	3	POR	End of 3rd	10:00p	http://ww	C. McColl	2015
18	35	28	39	28	7	11	4	28	28	14	14	42	2-pt	6	TRUE	6	1	3	POR	End of 1st	9:00p	http://ww	C. Kaman	2015

Figure: Analysis of the 2 for 1 strategy

Visualizing it is crucial!



Solving estimation problems

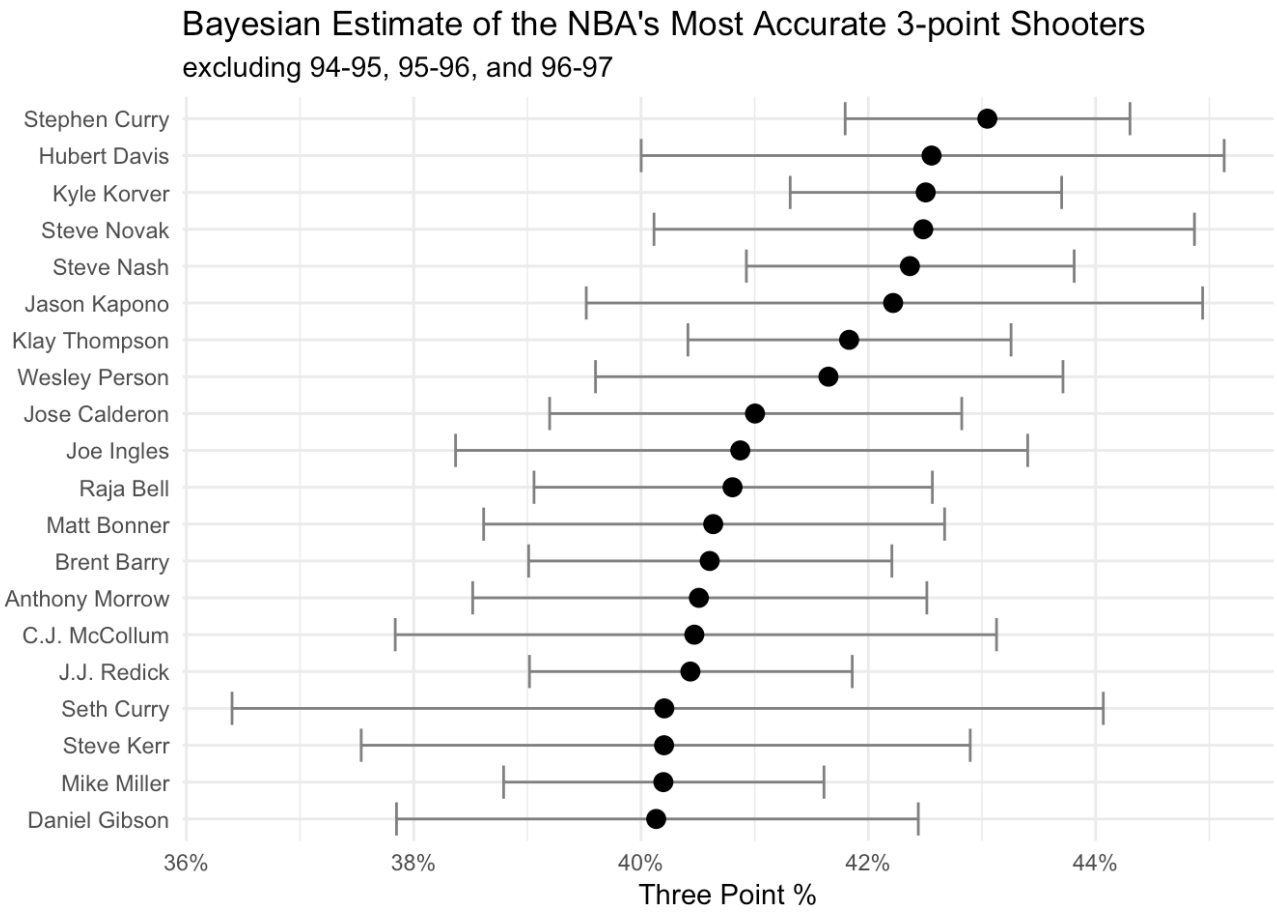


Figure: Estimation of the real 3-pt percentage of NBA players

What is the value of a player?

#	Player	Age	Nat.	Club	Market value	Matches								
1	Kylian Mbappé Right Winger	20			200,00 Mill. €	27	19	0	11	4	0	1	5	4
2	Neymar Left Winger	26			180,00 Mill. €	27	21	0	15	8	0	0	1	7
3	Lionel Messi Right Winger	31			160,00 Mill. €	21	22	0	13	1	0	0	2	1
4	Mohamed Salah Right Winger	26			150,00 Mill. €	30	17	0	8	1	0	0	3	9
5	Harry Kane Centre-Forward	25			150,00 Mill. €	35	21	0	9	5	0	0	4	5
6	Antoine Griezmann Centre-Forward	27			150,00 Mill. €	32	15	0	7	6	0	0	0	10
7	Kevin De Bruyne Attacking Midfield	27			150,00 Mill. €	11	3	0	2	1	0	0	5	6
8	Eden Hazard Left Winger	28			150,00 Mill. €	33	14	0	11	2	0	0	7	9
9	Philippe Coutinho Attacking Midfield	26			140,00 Mill. €	28	7	0	4	1	0	0	8	12
10	Raheem Sterling Right Winger	24			120,00 Mill. €	30	13	0	11	4	0	0	3	10
11	Paulo Dybala Second Striker	25			110,00 Mill. €	28	8	0	5	2	0	0	6	9
12	N'Golo Kanté Central Midfield	27			100,00 Mill. €	33	3	0	2	4	0	0	3	2
13	Dele Alli Attacking Midfield	22			100,00 Mill. €	26	6	0	5	2	0	0	4	8
14	Leroy Sané Left Winger	23			100,00 Mill. €	32	12	0	11	1	0	0	10	9
15	Mauro Icardi Centre-Forward	25			100,00 Mill. €	26	14	0	2	0	0	0	3	4
16	Cristiano Ronaldo Left Winger	33			100,00 Mill. €	25	15	0	8	2	0	1	2	1

Figure: TransferMarkt Estimation of soccer players

When should you start your coaching?

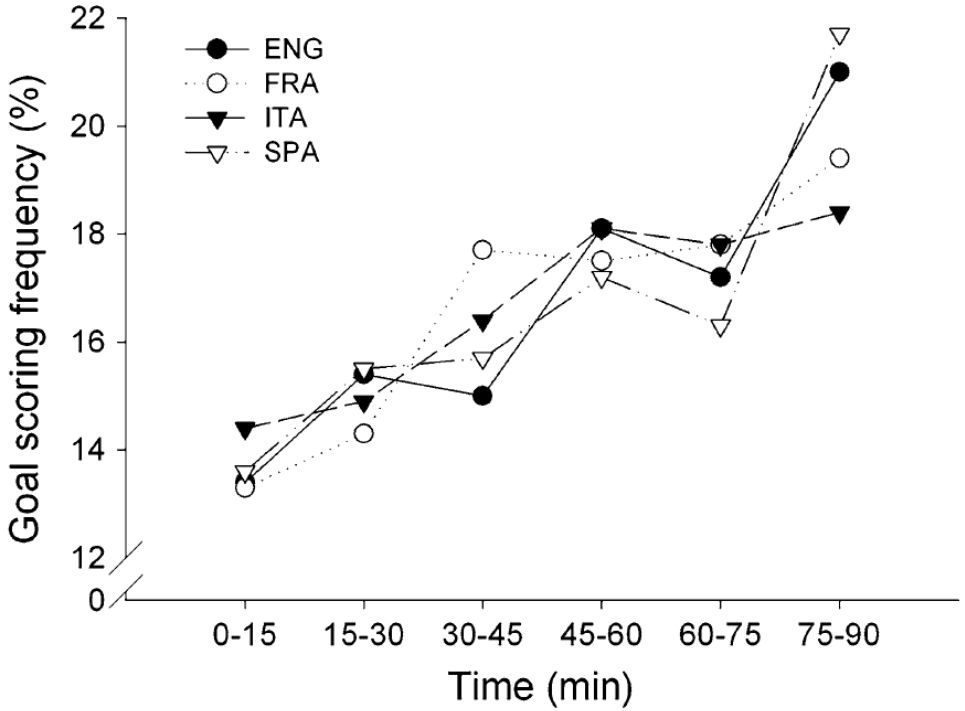


Fig. 2 The frequency of goals scored in each of the six 15-min intervals of the matches in each European league between the 2008–09 and 2010–11 competitive seasons (*ENG black circles* English premier league, *FRA white circles* French football Ligue 1, *ITA black triangles* Italian Serie A, *SPA white triangles* Spanish football Liga)

Can you predict injuries?

d_{TOT}	Distance in meters covered during the training session
d_{HSR}	Distance in meters covered above 5.5m/s
d_{MET}	Distance in meters covered at metabolic power
d_{HML}	Distance in meters covered by a player with a Metabolic Power is above 25.5W/Kg
$d_{HML/m}$	Distance in meters covered by a player with a Metabolic Power is above 25.5W/Kg per minute
d_{EXP}	Distance in meters covered above 25.5W/Kg and below 19.8Km/h
Acc_2	Number of accelerations above 2m/s2
Acc_3	Number of accelerations above 3m/s2
Dec_2	Number of decelerations above 2m/s2
Dec_3	Number of decelerations above 3m/s2
DSL	Total of the weighted impacts of magnitude above 2g. Impacts are collisions and step impacts during running
FI	Ratio between DSL and speed intensity
Age	age of players
BMI	Body Mass Index: ratio between weight (in kg) and the square of height (in meters)
Role	Role of the player
PI	Number of injuries of the players before each training session
Play time	Minutes of play in previous games
Games	Number of games played before each training session

https://doi.org/10.1371/journal.pone.0201264.t001

Can you even predict instantaneous plays?

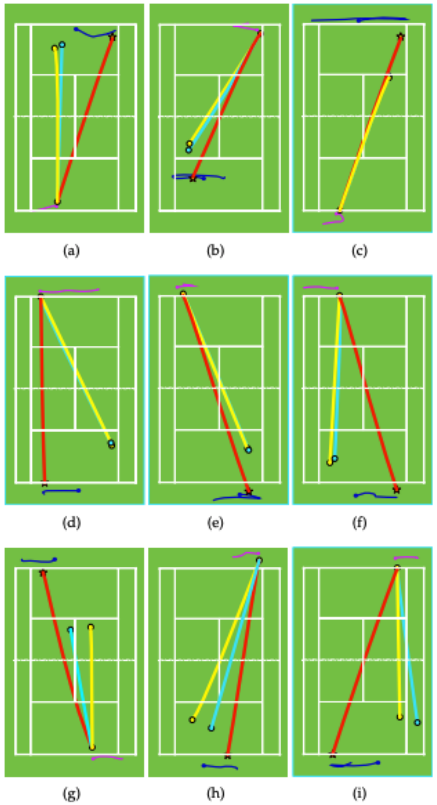
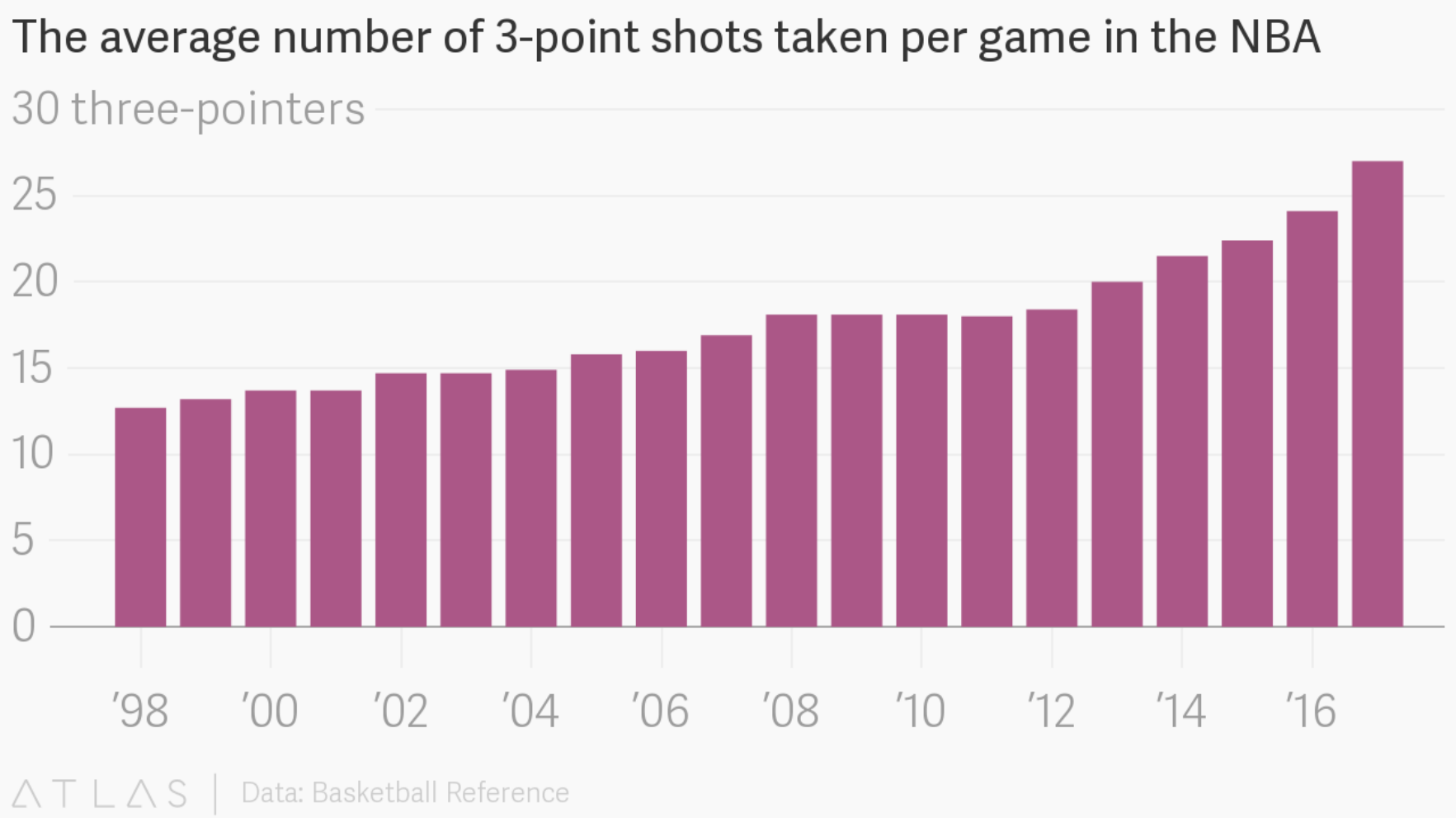


Fig. 5. Qualitative results from the proposed MSS-GAN model. Incoming shot is in red where the yellow star and the circle denotes the shot starting and ending locations. Feet movements of player of interest and the opponent are in magenta and blue colours. Ground truth and predicted trajectories are denoted in cyan and yellow lines, respectively.First two rows show accurate predictions while the 3rd row shows scenarios where the predicted trajectory deviates from the ground truths. However in (g) and (h) we observe that the predicted trajectory maximises the opportunity of the winning probability of the player of interest. Please note that we have overlaid the predictions from RGN on top of court outlines for the clarity of illustration

It has an impact!



References

Alberti, Giampietro & Iaia, F & Arcelli, Enrico & Cavaggioni, Luca & Rampinini, Ermanno. (2013).

Goal scoring patterns in major European soccer leagues.

Sport Sciences for Health. 9. 10.1007/s11332-013-0154-9.

Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernndez J, et al. (2018)

Effective injury forecasting in soccer with GPS training data and machine learning.

PLOS ONE 13(7): e0201264. https://doi.org/10.1371/journal.pone.0201264

Albert J.

Visual Count Effects

https://baseballwithr.wordpress.com/

Fernando T., Denman S., Sridharan S., Fookes C.

Memory Augmented Deep Generative models for Forecasting the Next Shot Location in Tennis arXiv:1901.05123 https://arxiv.org/abs/1901.05123

The End